International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Two Classes of Covariance Matrices Giving Simple Linear Forecasts

Jewell, W.S.

IIASA Research Memorandum
May 1975

RM-75-17

TWO CLASSES OF COVARIANCE MATRICES GIVING SIMPLE LINEAR FORECASTS

William S. Jewell

May 1975

ABSTRACT

Two special classes of covariance matrices are considered which give simplified computations for linear forecasts without continued reinversion of the matrix. In the first class, the optimal coefficients in the forecast can be computed in advance for every time period by simple closed formulas. In the second class, which is a generalization of the first, the optimal coefficients are obtained through a simple first-order linear recursive relation between forecasts of successive time periods. Collective risk forecasting models which give rise to these classes of covariances are presented.

TWO CLASSES OF COVARIANCE MATRICES GIVING SIMPLE LINEAR FORECASTS

William S. Jewell*

INTRODUCTION

Suppose we have a random vector, $\underline{\xi} = [\xi_1, \xi_2, \ldots, \xi_n]'$ from whose values $\underline{x} = [x_1, x_2, \ldots, x_n]$ we are trying to predict a random variable η through a *forecast function*, $f(\underline{x})$. Assuming that the joint distribution of $(\eta, \underline{\xi})$ is known, then the integrable function which minimizes the *mean-squared error*, $E\{(\eta - f(\underline{\xi}))^2\}$, is just the conditional mean, $f^O(\underline{x}) = E\{\eta | \underline{\xi} = \underline{x}\}$, sometimes called the *regression* of η on $\underline{\xi}$.

If this function is difficult to calculate, then a *linear regression*,

$$(1) \qquad f(\underline{x}) = a_O + \sum_{i=1}^{n} a_i x_i \quad ,$$

may be sought which makes the *approximation error*, $E\{(f^O(\underline{\xi}) - f(\underline{\xi}))^2\}$, as small as possible by adjusting the coefficients $a_O, a_1, \ldots, a_n$.

It is well known that the optimal values of these coefficients are given by a single equation which adjusts a_O to make the forecast unbiased,

$$(2) \qquad a_O = E\{\eta\} - \sum_{i=1}^{n} a_i E\{\xi_i\} \quad , \qquad E\{f(\underline{\xi})\} = E\{\eta\} \quad ,$$

*On leave from the Department of Operations Research, University of California, Berkeley.

together with an $n \times n$ system of linear equations for the remaining coefficients,

$$(3) \qquad \sum_{j=1}^{n} \text{Cov}\,\{\xi_i;\xi_j\} \cdot a_j = \text{Cov}\,\{\xi_i;\eta\} \quad , \quad (i = 1,2,\ldots,n).$$

Thus, the basic computational labor is in inverting the $n \times n$ covariance matrix,

$$(4) \qquad C_{ij} = \text{Cov}\,\{\xi_i;\xi_j\} \quad , \qquad (i,j = 1,2,\ldots,n),$$

and then premultiplying it into the RHS of (3).

In particular, if $\eta = \xi_{n+1}$, we are interested in *linear forecasts* for $n = 0,1,2,\ldots$, and the continued reinversion of matrices $C = [C_{ij}]$ of expanding order represents a formidable computational task in the general case. The fact that covariance matrices are positive (semi-)definite can lead to efficient iterative methods (see, e.g., [1]), but one would also like to have explicit or algorithmic exact solutions if at all possible.

The purpose of this paper is to present two special classes of covariance matrices which lead to simplified computation of (3) in the following sense:

(i) either an explicit solution for the $\{a_i\}$, and hence for $f(\underline{x})$, can be given for all i and all n;

(ii) or a recursive solution can be found for $f_{n+1}(x_1,x_2,\ldots,x_{n-1},x_n)$ in terms of $f_n(x_1,x_2,\ldots,x_{n-1})$ and the new data, x_n.

These classes of covariance matrices were suggested by recent results on collective risk models [4], [5], [6], [11].

Many of these results are not new, apparently being continually rediscovered in different fields of application. However, we feel that it is desirable to collect in one place all known results which may be useful in linear regression problems, and to show how these matrices arise naturally in various prediction problems in collective risk theory.

Without loss of generality, we decompose the covariance matrix as follows:

$$(5) \qquad C_{ij} = \begin{cases} E_{ii} + D_{ii} & (i = j) \\ \\ D_{ij} & (i \neq j) \end{cases} ,$$

and note that in forecasting problems, $\mathrm{Cov}\,\{\xi_i;\eta\} = D_{i,n+1}$. Our basic system (3) now reads:

$$(6) \qquad E_{ii} \cdot a_i + \sum_{j=1}^{n} D_{ij} \cdot a_j = D_{i,n+1} \quad , \quad (i = 1,2,\ldots,n),$$

where it is important to note that all coefficients $a_1, a_2, \ldots, a_n$ are now included in the sum. In the problems of practical interest, no E_{ii} is zero, so that (6) can be written in an obvious matrix notation:

$$(7) \qquad (I_n + \Delta)\, \underline{a} = \Delta_{(n+1)} \quad ,$$

where the $n \times n$ matrix Δ and the n-vector $\Delta_{(n+1)}$ have coefficients

(8) $$\Delta_{ij} = D_{ij}/E_{ii} \quad ,$$

and I_n is the unit matrix of order n. From this, it is clear that simplified computation depends on a special form for the off-diagonal elements of D_{ij}.

MODEL I: EXPLICIT SOLUTION

Basic Result

The first model assumes that D_{ij} may be factored into

(9) $$D_{ij} = \alpha_i \cdot \beta_j \quad .$$

Substituting into (6) gives an explicit solution for a_i,

(10) $$a_i = \alpha_i \left(\beta_{n+1} - \sum_{j=1}^{n} \beta_j a_j \right) \Big/ E_{ii} \quad , \qquad (i = 1,2,\ldots,n),$$

in terms of an unknown sum which is the same for all i. But this sum β can be found explicitly by performing the indicated sum,

$$\beta = \sum_{j=1}^{n} \beta_j a_j = (\beta_{n+1} - \beta) \sum_{j=1}^{n} (\alpha_j \beta_j / E_{jj}) \quad ,$$

or,

$$\beta_{n+1} - \beta = \beta_{n+1} \left(1 + \sum_{j=1}^{n} (\alpha_j \beta_j / E_{jj}) \right)^{-1} \quad .$$

Substituting back in (10), we have finally the explicit solution:

$$(11) \qquad a_i = \frac{\alpha_i \beta_{n+1}}{E_{ii}\left(1 + \sum_{j=1}^{n} (\alpha_j \beta_j / E_{jj})\right)} = \frac{(D_{i,n+1}/E_{ii})}{1 + \sum_{j=1}^{n} (D_{jj}/E_{jj})} ,$$

for $i = 1,2,\ldots,n$, with a_0 given by (2).

Related Results

The above result is related to the following:

Theorem. If α and β are $n \times k$ matrices, then

$$(12) \qquad (I_n + \alpha\beta')^{-1} = I_n - \alpha(I_k + \beta'\alpha)^{-1} \beta' ,$$

whenever either of the indicated inverses exists.

Bodewig ([1], pp. 39, 218) attributes this result for $k = 1$ to Bartlett, and the generalization to Hemes. The general result is also given by Tocher [17], and two later attributions may be found in [15], pp. 6, 34. The fact that the determinant of the two terms in parenthesis in (12) are equal ([19], p. 231) shows that the existence of one inverse implies the existence of the other.

(11) now follows directly from the fact that Δ is a simple product ($k = 1$). The general result is used in (17) below.

Note that (12) does not require Δ to be symmetric. If we add the fact that $D_{ij} = D_{ji}$, then α_i and β_i can only differ by a constant multiplier D_{00}, so that $D_{ij} = \alpha_i D_{00} \alpha_j$, or if the constant is absorbed equally, (9) may be replaced by $D_{ij} = \alpha_i \alpha_j$.

Matrix Generalization

The same idea can be used to reduce computational labor in problems of higher dimension. For example, suppose that the $\underline{\xi}_i$ are themselves row vectors $\underline{\xi}_i = [\xi_{i1}, \xi_{i2}, \ldots, \xi_{ip}]$, so that n samples generate a data matrix $X = \{x_{ik};\ i = 1,2,\ldots,n,\ k = 1,2,\ldots,p\}$. Then the coefficients in the linear estimator of, say, the s^{th} component of $\underline{\xi}_{n+1}$,

$$f_0(X) = a_{0s} + \sum_{i=1}^{n} \sum_{k=1}^{p} a_{iks} x_{ik} \quad , \tag{13}$$

will be given by (n+p) equations of the form:

$$\sum_{j=1}^{n} \sum_{\ell=1}^{p} \mathrm{Cov}\ \{\xi_{ik};\xi_{j\ell}\}\ a_{j\ell s} = \mathrm{Cov}\ \{\xi_{ik};\xi_{n+1,s}\} \quad , \tag{14}$$

for i = 1,2,...,n and k = 1,2,...,p, together with a single equation similar to (2) to make the forecast unbiased.

Again, without loss of generality, we write

$$\mathrm{Cov}\ \{\xi_{ik};\xi_{j\ell}\} = \begin{cases} E_{ik,i\ell} + D_{ik,i\ell} & (i = j) \\ \\ D_{ik,j\ell} & (i \neq j) \end{cases} \tag{15}$$

as being convenient for collective risk models, and imagine that both of these coefficient sets are grouped into $n \times n$ arrays of square submatrices, $E(i,i)$ and $D(i,j)$, each of which is of size $p \times p$. Thus, $[D(i,j)]_{k,\ell} = D_{ik,j\ell}$.

The coefficients $a_{j\ell s}$ and the RHS of (14) can also be partitioned into n vector blocks of p coefficients each, since s and n+1 are fixed for this forecast. However, it is notationally more convenient if we imagine the RHS of (14) being augmented by *all* the columns $s = 1,2,\ldots,p$; then the coefficients $a_{j\ell s}$ can be represented by n blocks of square $p \times p$ submatrices, call them A(i), and the RHS becomes blocks of matrices D(i,n+1). In effect, the solution to this expanded system will give *all* the coefficients for any possible component prediction at the same time. ((13) could also be written in matrix format; see [10]).

In this block matrix notation then, (14) becomes:

$$(16) \qquad E(i,i)\ A(i) + \sum_{j=1}^{n} D(i,j)\ A(j) = D(i,n+1) \quad .$$

The simplification corresponding to (9) assumes that each of the $p \times p$ submatrices D(i,j) factors into a matrix product of two similar matrices, $\alpha(i)$, $\beta(j)$:

$$(17) \qquad D(i,j) = \alpha(i) \cdot \beta(j) \quad , \qquad (i,j = 1,2,\ldots,n).$$

The solution procedure is similar to the scalar case, with the final result:

$$(18) \quad A(i) = E^{-1}(i,i)\ \alpha(i) \left[I_p + \sum_{j=1}^{n} \beta(j)\ E^{-1}(j,j)\ \alpha(j) \right]^{-1} \beta(n+1)$$

for $i = 1,2,\ldots,n$, which should be compared with (11), (12). We see that the computational labor has been reduced from inverting the (n+p) × (n+p) covariance matrix to that of inverting n submatrices $E(i,i)$ of order $p \times p$, performing some multiplications and summations, followed by one more inversion of order $p \times p$ to find the factor common to all $A(i)$.

It is difficult to get (18) into a form similar to the second equation of (11), as can be seen from the rearrangements:

(19)

$$A(i) = E^{-1}(i,i)\left[I_p + \sum_{j=1}^{n} D(i,j)\, E^{-1}(j,j)\, D(j,j)\, D^{-1}(i,j)\right]^{-1} D(i,n+1)$$

$$= E^{-1}(i,i)\, D(i,n+1)\left[I_p + \sum_{j=1}^{n} D^{-1}(j,n+1)\, D(j,j)\, E^{-1}(j,j)\, D(j,n+1)\right]^{-1} .$$

Unless the matrices have special forms, the first and last terms in the sums do not cancel out, as in the scalar case.

Perhaps the easiest computational sequence is to work directly with a reduced $p \times p$ matrix, $e(i,i)$, calculated for each i by:

(20) $$e^{-1}(i,i) = \beta(i)\, E^{-1}(i,i)\, \alpha(i) \quad ,$$

and then obtain a matrix of reduced coefficients

(21) $$\beta(i)\, A(i)\, \beta^{-1}(n+1) = e^{-1}(i,i)\left[I_p + \sum_{j=1}^{n} e^{-1}(j,j)\right]^{-1} .$$

This can be re-inflated for direct use, or one may rearrange the vector form of (13) in terms of the reduced coefficients and reduced data [11].

The above result does not use the known symmetry of the covariance matrix, which implies that $D(i,j) = D'(j,i)$. Then $\alpha(i)$ and $\beta'(j)$ can only differ by a constant symmetric matrix D_{00}, so that $D(i,j) = \alpha(i)\, D_{00}\, \alpha'(j)$, or if D_{00} has a square root, it may be absorbed into the definition of $\alpha(i)$, giving $D(i,j) = \alpha(i)\, \alpha'(j)$.

Applications in Collective Risk Forecasting

In the model of collective risk forecasting used extensively in casualty insurance, we imagine that each random variable $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}$ depends upon a fixed, but unknown, *risk parameter* θ. Furthermore, given θ, the samples $x_1, x_2, \ldots, x_n$ are independent. The problem is then to predict the mean value of the next sample, ξ_{n+1}, given the data, or, in insurance terminology, to find the *fair premium* for period n+1, given the *experience data* on a single risk, and *collective statistics* for other risks with differing risk parameters [3].

If we imagine that these statistics are available as a *prior density* on θ, $p(\theta)$, and a *likelihood* (conditional density) $p_i(x_i|\theta)$ for each ξ_i $(i = 1,2,\ldots,n,n+1)$, then the forecast problem can be seen to be equivalent to a *Bayesian forecast* of the conditional mean [8], [14]. If we further require that the forecast be linear in the data, then we have a linearized Bayesian forecast, which is called a *credibility forecast* in actuarial literature. This is nothing more than a linear regression (1), (2), (3), with a special form for the covariance matrix reflecting the prior collective mixture of different risk parameters.

Using the prior and the likelihood, we see that the required first and second moments are:

$$(22) \quad m_i = E\{\xi_i\} = E_\theta m_i(\theta) \quad ; \quad m_i(\theta) = E\{\xi_i|\theta\} = \int x_i p_i(x_i|\theta)\, dx_i \quad ;$$

$$(23) \quad C_{ij} = \mathrm{Cov}\,\{\xi_i;\xi_j\} = E_\theta\, \mathrm{Cov}\,\{\xi_i;\xi_j|\theta\} + \mathrm{Cov}_\theta\,\{m_i(\theta);m_j(\theta)\} \quad .$$

However, because of the independence of the samples, given θ, the first term of (23) is nonzero only for $i = j$, and the definitions:

$$(24) \quad E_{ii} = E_\theta\, \mathrm{Var}\,\{\xi_i|\theta\} = \iint (x_i - m_i(\theta))^2\, p_i(x_i|\theta)\, p(\theta)\, dx_i d\theta \quad ;$$

$$(25) \quad D_{ij} = \mathrm{Cov}_\theta\,\{m_i(\theta);m_j(\theta)\} = \int (m_i(\theta) - m_i)(m_j(\theta) - m_j)\, p(\theta)\, d\theta \quad ,$$

are consistent with (5). The first group is called the *mean variances*, and the second group the *covariance of the means*.

In classical credibility theory (see, e.g., [3]) the ξ_i are identically distributed over the samples, so that the only collective statistics needed are the common values, m, E, D. Then the solution of (6) is easily:

$$(26) \qquad a_i = D/(E + nD) \quad , \qquad (i = 1,2,\ldots,n),$$

giving the forecast:

$$(27) \qquad f(\underline{x}) = (1 - Z)\, m + Z\left(\sum_{i=1}^{n} x_i/n\right)$$

with *credibility factor*:

$$(28) \qquad Z = n/(n + (E/D)) \quad .$$

There are many interesting aspects to this result, one of which is that as $n \to \infty$, the credibility attached to the sample mean approaches unity. There are vector forms of this result [10], [11], and for certain families of priors and likelihoods, it can be shown to give an exact forecast of the conditional mean [8], [9], [10].

Turning to time-varying models, Bühlmann and Straub [4], [5] have investigated a class of models in which the total losses on an insurance contract in period i are normalized by dividing by V_i, the volume, or *exposure*, of business in the same period. ξ_i is then the loss on a per-unit basis, which leads to:

(29) $$m_i(\theta) = m_0(\theta) \quad ; \quad D_{ij} = D_{00} = \mathrm{Var}_\theta \{m_0(\theta)\} \quad ;$$

(30) $$E_{ii} = E_{00}/V_i \quad ,$$

where D_{00} and E_{00} are the estimated unit exposure values for variance of the mean, and mean variance over the collective.

In terms of simplification (9), this model has $\alpha_i = 1$, $\beta_j = D_{00}$ for all i,j, giving, finally,

(31) $$f(\underline{x}) = (1 - Z)\, m_0 + Z\left(\sum_{i=1}^{n} V_i x_i \Big/ \sum_{j=1}^{n} V_j\right)$$

with credibility factor

(32) $$Z = \sum_{j=1}^{n} V_j \Big/ \left(\sum_{k=1}^{n} V_k + (E_{00}/D_{00})\right) \quad .$$

This can be seen to be similar to (27), (28), except that the "operational time" is now measured in volume units. Bühlmann and Straub also consider many other related models in which the separability of D_{ij} leads to closed forms.

In [11], the author considers a one-dimensional, time-varying model with *separable mean*, in which it is assumed that the known dependency of the mean risk over time can be factored out, as:

(33) $$m_i(\theta) = \alpha_i m_O(\theta) \quad ,$$

giving

(34) $$D_{ij} = D_{OO}\alpha_i\alpha_j \quad ; \quad D_{OO} = \mathrm{Var}_\theta \{m_O(\theta)\}$$

for all (i,j). The mean variances, E_{ii}, remain arbitrary; thus the correspondence with our previous notation is immediate, and we obtain either (11), or, in more revealing format:

(35) $$f(\underline{x}) = \alpha_{n+1}\left[\left(1 - \sum_{i=1}^{n} z_i\right) m_O + \sum_{i=1}^{n} z_i\left(\frac{x_i}{\alpha_i}\right)\right] \quad ,$$

where $m_O = E_\theta m_O(\theta)$, and the *per-observation credibility factors*, z_i, are:

(36) $$z_i = (D_{OO}/E_{ii})\Big/\left(1 + \sum_{j=1}^{n} (D_{OO}/E_{jj})\right) \quad ,$$

for i = 1,2,...,n. In other words, each observation is normalized by the factor α_i, weighted by z_i which depends only on the ratios D_{00}/E_{ii}, and then "re-inflated" to period n+1 by the factor α_{n+1}. The use of reciprocal variances as weights is well known in statistics for observations with normal distributions of error.

In a later section of [11], the author also treats the multidimensional separable mean, in which it is assumed (in current notation) that

$$(37) \qquad m_{ik}(\theta) = \alpha_{ik} m_{0k}(\theta) \quad ,$$

so that

$$(38) \quad D_{ik,j\ell} = \alpha_{ik} D_{0k,0\ell} \alpha_{j\ell} \;\; ; \;\; D_{0k,0\ell} = \mathrm{Cov}_\theta \, \{m_{0k}(\theta); m_{0\ell}(\theta)\} \quad .$$

In the matrix notation of the previous section, this makes $D_{0k,0\ell} = [D_{00}]_{k\ell}$, and $\alpha(i) = \mathrm{Diag}\,\{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ip}\}$. The solution has coefficients similar to (21), but will not be reproduced here.

MODEL II: RECURSIVE SOLUTION

General Remarks

In the event that the optimal coefficients a_i cannot be found explicitly, a computational simplification still results if they can be found *recursively* for n = 1,2,3,...; this is especially desirable in forecasting problems, where previous predictions are available for use with the current value of n. Henceforth, let $a_i(n)$ refer to the coefficients used to predict ξ_{n+1} in the forecast function $f_{n+1} = f(x_1, x_2, \ldots, x_n)$; the covariance matrix at this stage of the computation will be called C(n), and is of order n × n.

For general C(n), there are explicit matrix formulas available for updating, based upon a relation due to Frobenius-Schur. First, partition C(n) as follows:

$$C(n) = \left[\begin{array}{c|c} C(n-1) & \underline{u} \\ \hline \underline{u}' & C_{nn} \end{array}\right] ,$$

where the (n-1)-vector $\underline{u} = [C_{n1}, C_{n2}, \ldots, C_{n,n-1}]'$, and we use the fact that all C's are symmetric. The Frobenius-Schur inverse of C(n) is then [1]:

$$(39) \quad C^{-1}(n) = \left[\begin{array}{c|c} C^{-1}(n-1) & \underline{O} \\ \hline \underline{O}' & O \end{array}\right] + k^{-1} \left[\begin{array}{c|c} \underline{vv}' & -\underline{v} \\ \hline -\underline{v}' & 1 \end{array}\right] ,$$

where $\underline{O}$ is an (n-1)-vector of zeroes,

$$(40) \qquad \underline{v} = C^{-1}(n-1)\ \underline{u} = [a_1(n-1), a_2(n-1), \ldots, a_{n-1}(n-1)]' \quad ,$$

and

$$(41) \qquad k = C_{nn} - \underline{u}'\ C^{-1}(n-1)\ \underline{u} \quad .$$

Thus, successive inverses of C can be found in an efficient way from the previous inverses, starting with $C^{-1}(1) = [C_{11}{}^{-1}]$. At each step, the optimal forecast coefficients are then obtained by multiplying $C^{-1}(n)$ into the $\underline{u}$ for the $(n+1)^{st}$ problem. This useful relation is continually being rediscovered in a variety of applications of the least-squares method.

However, for our purposes, it is still too complex, since an inverse of increasing size must be stored, and matrix operations continue to be required when only a single answer is sought.

By examining successive ratios of the coefficients (11) for problems of different sizes, we see that a simple recursive solution for Model I is given by:

$$(42)\quad \frac{a_i(n)}{a_i(n-1)} = \left(\frac{\beta_{n+1}}{\beta_n}\right)\left\{1 + (D_{nn}/E_{nn})\left[1 + \sum_{j=1}^{n-1} (D_{jj}/E_{jj})\right]^{-1}\right\}^{-1}$$

for *every* $i = 1,2,\ldots,n-1$, and $n = 2,3,\ldots$. The boundary value coefficients are:

$$(43)\quad a_n(n) = (D_{n,n+1}/E_{nn}) \Big/ \left[1 + \sum_{j=1}^{n} (D_{jj}/E_{jj})\right], \quad (n = 1,2,\ldots)$$

since $a_i(n) \equiv 0$, $(i = n+1, n+2, \ldots)$. ($a_0(n)$ is given by (2)).

Even this computation can be further simplified by defining a secondary recursive sequence $\{b_n\}$, as follows:

$$(44)\qquad b(n) = b(n-1) + (D_{nn}/E_{nn}) \quad ; \quad b(0) = 1 \quad ,$$

and noting that the forecast functions $f_1, f_2, \ldots$ can be written recursively as:

$$(45)\quad f_{n+1} = \left(\frac{\beta_{n+1}}{\beta_n}\right)\left(\frac{b_{n-1}}{b_n}\right) f_n + \left(\frac{D_{n,n+1}}{b_n E_{nn}}\right) x_n + \left[m_{n+1} - \left(\frac{\beta_{n+1}}{\beta_n}\right) m_n\right] \quad ,$$

$$(n = 1,2,\ldots),$$

$$f_1 = m_1 \quad .$$

This clearly simplifies storage and computation for Model I, since only the most recent values of $b(n)$ and f_n need to be retained.

We shall now examine what more general forms for D_{ij} lead to first-order linear recursion relationships similar to (45). This work was motivated by a paper of Gerber and Jones [6].

First-Order Linear Recursion

Temporarily, let us simplify the algebra by assuming that the means of all observations have been normalized to the same value, m_1. Data of this type is said to be in "as-if" form.

Assume that there are known sequences $(\pi_1, \pi_2, \ldots)$, $(\rho_1, \rho_2, \ldots)$ such that the forecast functions $f_1, f_2, \ldots$ follow a first-order linear recursive relationship:

(46) $$f_{n+1} = \pi_n f_n + \rho_n x_n + (1 - \pi_n - \rho_n)\, m_1 \quad , \qquad (n = 1,2,\ldots),$$

$$f_1 = m_1 \quad .$$

Note that in this form the forecast is unbiased. Now, what form of the D_{ij} could lead to this result?

First, (46) implies:

(47) $$a_n(n) = \rho_n \quad , \qquad (n = 1,2,\ldots),$$

(48) $$a_i(n) = \pi_n a_i(n-1) \quad , \qquad (i = 1,2,\ldots,n-1).$$

Then, if two versions of (6) are written for $a_i(n)$ and $a_i(n-1)$, for $i < n$, the use of (47), (48) leads to:

(49) $$D_{i,n+1} = (\pi_n + \rho_n)\, D_{in} \quad , \qquad (i = 1,2,\ldots,n-1),$$

which must hold for $n = 2,3,\ldots,$ so that

(50) $$D_{ij} = D_{ii} \prod_{k=i}^{j-1} (\pi_k + \rho_k) \quad , \qquad (i < j).$$

Now, using the fact that D_{ij} is symmetric, we see that the general form for all i,j must be expressable as:

(51) $$D_{ij} = \lambda_{\mathrm{Min}(i,j)} \cdot \mu_{\mathrm{Max}(i,j)} \quad ,$$

with, as one possible choice:

(52) $$\lambda_i = D_{ii}/\mu_i \quad , \qquad (i = 1,2,\ldots),$$

and

(53) $$\mu_i = \prod_{k=1}^{i-1} (\pi_k + \rho_k) \quad , \quad (i = 2,3,\ldots) \quad ; \quad \mu_1 = 1.$$

The diagonal elements, $\{D_{ii}, E_{ii}\}$ are related to $\{\pi_n, \rho_n\}$ through a recursion relation which is gotten from the $(n-1)^{st}$ and n^{th} equations of (6) for $a_i(n-1)$ and $a_i(n)$, respectively. We get:

(54)

$$(E_{nn} + D_{nn})\, \rho_n = (\pi_n E_{n-1,n-1})\, \rho_{n-1} + D_{n,n+1} - \pi_n D_{n-1,n} \quad , \quad (n = 2,3,\ldots),$$

$$(E_{11} + D_{11})\, \rho_1 = D_{12} \quad ,$$

which can be manipulated in a variety of ways, depending upon what data are given. For example, if the weightings $\{\pi_n, \rho_n\}$ are given for all n, then the diagonal covariance elements are related through:

(55)

$$E_{nn} = \left(\frac{\pi_n}{\rho_n}\right)\left[\rho_{n-1}E_{n-1,n-1} + D_{nn} - (\pi_{n-1} + \rho_{n-1})\, D_{n-1,n-1}\right] , \quad (n = 2,3,\ldots),$$

$$E_{11} = (\pi_1/\rho_1)\, D_{11} \quad .$$

Alternatively, if the D_{ij} are given, and observed to be in form (51), then from the factors D_{ii}, E_{ii} and μ_i, we can calculate the factors π_n and ρ_n as follows:

(56)

$$\rho_n = \left(\frac{\mu_{n+1}}{\mu_n}\right)\left\{1 + E_{nn}\left[D_{nn} - \left(\frac{\mu_n}{\mu_{n-1}}\right) D_{n-1,n-1} + \rho_{n-1}E_{n-1,n-1}\right]^{-1}\right\}^{-1} ,$$

$$(n = 2,3,\ldots),$$

$$\rho_1 = \mu_2 D_{11}/(D_{11} + E_{11}) \quad .$$

The factor $D_{nn} - (\mu_n/\mu_{n-1})\, D_{n-1,n-1}$ is, of course, $\lambda_n(\mu_n - \mu_{n-1})$.

Following Gerber and Jones [6], we note that (56) can be simplified through a new recursive sequence $\{U_i\}$ such that:

(57)

$$U_n = D_{nn} - (\mu_n/\mu_{n-1})\, D_{n-1,n-1} + (\mu_n/\mu_{n-1})\left[E^{-1}_{n-1,n-1} + U^{-1}_{n-1,n-1}\right]^{-1} ,$$

$$U_1 = D_{11} \quad ,$$

giving

(58) $$\rho_n = (\mu_{n+1}/\mu_n)\,\frac{U_n}{E_{nn} + U_n} \quad , \qquad (n = 1,2,\ldots).$$

The factors π_n are then simply:

(59) $$\pi_n = (\mu_{n+1}/\mu_n) - \rho_n \quad ,$$

for all n, remembering that $\mu_1 = 1$.

Once the $\{\pi_n, \rho_n\}$ are calculated, the optimal weighting coefficients at the n^{th} step follow directly from the definition (46):

(60) $$a_0(n) = m_1\left[1 - \sum_{j=1}^{n} a_j(n)\right] \quad ;$$

$$a_j(n) = \rho_j \pi_{j+1} \pi_{j+2} \cdots \pi_n \quad ; \qquad (j = 1,2,\ldots,n-1);$$

$$a_n(n) = \rho_n \quad .$$

Now let us reconsider what happens if the means, $m_1, m_2, \ldots, m_n, m_{n+1}$, are in fact different from one another. By normalizing the variables to unity mean, $\xi_i^* = \xi_i/m_i$, we see that the above theory is applicable to the covariance components $D_{ij}^* = D_{ij}/m_i m_j$ and $E_{ii}^* = E_{ii}/m_i^2$. After some algebra, we find from (54) that, instead of the forecast (46), we obtain the result:

(61)

$$f_{n+1} = \left(\frac{m_{n-1}}{m_n}\right) \pi_n^0 f_n + \rho_n^0 x_n + (m_{n+1} - m_n\rho_n^0 - m_{n-1}\pi_n^0) \quad , \quad (n = 2,3,\ldots),$$

$$f_1 = m_1 \quad ,$$

where $\{\pi_n^0, \rho_n^0\}$ are the coefficients that would be obtained from the previous theory (47)-(60) by using the *same* E_{ii} and D_{ij}, but *neglecting* the difference in the $\{m_i\}$. Note particularly that the changing mean is compensated for in the new π_n and the constant term, but that the new-data multiplier, $\rho_n = \rho_n^0$, remains the same.

Related Results

In [6], Gerber and Jones investigated the "credibility" forecast form:

(62)
$$f_{n+1} = (1 - \rho_n) f_n + \rho_n x_n \quad , \quad (n = 2,3,\ldots),$$

$$f_1 = m_1 \quad ,$$

for constant mean, and thus obtained matrices of the form $D_{ij} = \lambda_{Min(i,j)}$. Since their development was followed in the generalization (47)-(60), their results can be gotten by setting $\mu_j = 1$ and $\pi_j = 1 - \rho_j$ for all j.

The matrix (51) is essentially the same as one analyzed by Roy and Sarhan [16] (see also [1], p. 222):

$$D_{ij} = c_i c_j (d_1 + d_2 + \dots + d_{Min(i,j)}) \quad , \tag{63}$$

where the $\{c_i, d_i\}$ are given constants easily related to $\{\lambda_i, \mu_i\}$. They show that D_{ij} has the triangular decomposition:

$$D = L\, D^{O}\, L' \quad , \tag{64}$$

with, in our notation:

$$L_{ij} = \begin{cases} O & (i < j) \\ \mu_i & (i \geq j) \end{cases} \tag{65}$$

and

$$D^{O} = \mathrm{Diag} \left\{ \frac{D_{11}}{\mu_1^2} \; ; \; \frac{D_{22}}{\mu_2^2} - \frac{D_{11}}{\mu_1^2} \; ; \; \frac{D_{33}}{\mu_3^2} - \frac{D_{22}}{\mu_2^2} \; ; \; \dots \right\} \quad . \tag{66}$$

From this, it follows that D^{-1}, and thus Δ^{-1}, are tridiagonal in form, so that efficient methods of computing the inverse (7) are possible.

In its continuous integral-operator form, (51) is the covariance of the so-called Gauss-Markov processes, which are used extensively in modelling communication detection and estimation problems, as well as control and regulation problems [12], [13]. The typical optimal prediction problem leads to a continuous operator version of (7), a Fredholm integral equation of the second kind. The recognized importance of the form (51) is that a factorization similar to (64) is possible, and this leads to simplified computations via a nonlinear Ricatti differential equation, whose properties have been extensively investigated. (I would like to thank J. Casti for these references.)

Generalizations

A natural generalization of (46) is to permit f_{n+1} to be predicted by a K^{th}-order recursion relationship, using $f_n, f_{n-1}, \ldots, f_{n-K+1}$, and x_n. This leads to a generalized version of (49), which links together K+1 successive D_{ij} in the same row, and to more complicated versions of (54), linking together the otherwise arbitrary E_{ii} and D_{ij}, for $(j \leq i+K-1)$. Although these results are easy to obtain, they are not particularly instructive in the absence of a model which might generate these forms. Electrical engineers, however, would be interested in such "realizable filters" as approximations to theoretically exact predictors. More complicated, but usually stationary, predictive models are used in the ARIMA forms of time series analysis [2].

In another direction, one can develop a matrix generalization of (46) similar to that of Model I. This would be natural for multidimensional problems which might have a simple covariance of means as between time periods, but not between different dimensions in the same time period. Further details are left to the reader.

Application in Collective Risk Forecasting

To illustrate how a collective risk model can lead to the form (51) and a forecast (61), we generalize an *evolutionary model* [11] due to Gerber and Jones [6]. (See also [18].)

In contrast to the previous assumption of a fixed risk parameter θ, we now assume that the parameter for a given sample is allowed to change over time according to a known law, giving $\theta_1, \theta_2, \ldots, \theta_n, \theta_{n+1}, \ldots$; the likelihood, given θ, may or may not change. Specifically, we suppose that the evolutionary mechanism provides a sequence of *mutually independent scale and location shifts* $\{\kappa_i, \sigma_i\}$ to the location parameters, $\{m_i(\theta_i)\}$, of the $\{\xi_i\}$, so that:

(67) $$m_i(\theta_i) = \kappa_i m_{i-1}(\theta_{i-1}) + \sigma_i \quad , \qquad (i = 1,2,\ldots),$$

and θ_{i-1} and $\{\kappa_i, \sigma_i; \kappa_{i+1}, \sigma_{i+1}; \ldots\}$ are mutually independent.

Further, assume that the first two moments of the shifts are given:

(68) $$E\{\kappa_i\} = k_i \neq 0 \quad ; \quad \text{Var}\,\{\kappa_i\} = G_i \geq 0 \quad ;$$

(69) $$E\{\sigma_i\} = s_i \quad ; \quad \text{Var}\,\{\sigma_i\} = H_i \geq 0 \quad ,$$

for $i = 1,2,\ldots$. It follows easily from the definitions that:

(70) $$m_i = k_i m_{i-1} + s_i$$
$$= \left(\prod_{j=2}^{i} k_j\right) m_1 + \sum_{j=2}^{i} s_j \prod_{\ell=2}^{j-1} k_\ell \quad ,$$

and

(71) $$D_{ii} = (G_i + k_i^2)\, D_{i-1,i-1} + (G_i m_i^2 + H_i)$$

$$= \left(\prod_{j=2}^{i} (G_j + k_j^2)\right) D_{11} + \sum_{j=2}^{i} (G_j + m_j^2) \prod_{\ell=2}^{j-1} (G_\ell + k_\ell^2) \quad ,$$

where the last product in both formulas is to be interpreted as unity when $j = 2$. More importantly, the general term for the covariance of the means is:

(72) $$D_{ij} = \frac{k_1, k_2, \ldots, k_{Max(i,j)}}{k_1, k_2, \ldots, k_{Min(i,j)}} D_{Min(i,j),Min(i,j)} \quad ,$$

so that the problem is of form (51).

Note specifically that it was *not* assumed that $\xi_i = \kappa_i \xi_{i-1} + \sigma_i$, given θ_i, so that the mean variances E_{ii} may vary in any desired manner.

In the Gerber and Jones model [6], $s_i = H_i = 0$, $G_i = G$ and $k_i = 1$ for all i, which leads to the simpler matrix form described earlier.

Gerber and Jones are also interested in special models which lead to geometric weights, instead of the usual credibility form (26). From (60) we see that $a_j(n) = \rho\pi^{n-j}$, $(j = 1,2,\ldots,n)$, when $D_{ij} = D_{Min(i,j),Min(i,j)} (\pi + \rho)^{|j-i|}$. Successive E_{ii} and D_{ii} must satisfy a relation similar to (55). If, in addition, we require that $E_{ii} = E$ for all i, it follows that $D_{ii} = (\pi + \rho)^{i-1} D_{11}$ in order to obtain geometric weights for all n. Finally, many families of (λ_i, μ_i) are asymptotically geometric, when (56) and (59) have stable fixed-point solutions [6].

A surprising result is obtained if we take $D_{ij} = D_{Min(i,j),Min(i,j)}$ and $E_{ii} - E_{i-1,i-1} = (i-1)(D_{ii} - D_{i-1,i-1})$ for all i,j. Then we find $\rho_n = 1 - \pi_n = n^{-1}$, and obtain the forecasts:

$$f_{n+1}(\underline{x}) = (x_1 + x_2 + \cdots + x_n)/n \quad , \qquad (n = 1,2,\ldots) \tag{73}$$

$$f_1 = m_1 \quad .$$

In insurance terminology, this forecast is "fully credible", because once the sample data become available, only it is used, and nothing about the collective need be known.

CONCLUSION

To summarize, the first class of covariance matrices, whose off-diagonal elements are $D_{ij} = \alpha_i D_{00} \alpha_j$, is included in the second class, whose elements are $D_{ij} = \lambda_{Min(i,j)} \mu_{Max(i,j)}$. However, the first class has the advantage that the optimal forecast coefficients can be computed once and for all, for all n; furthermore, the essential simplification is a property only of the covariance matrix, and thus will apply also to more general regression problems. The second class of covariance matrices uses explicitly the symmetry property, and the fact that the forecast RHS is a portion of the new column for the covariance matrix of higher order; this leads to a simple recursion relationship between forecasts in successive time periods.

Perhaps in this era of rapid digital computation, there is little need to stress computational simplicity of certain models. However, one is always interested in comparing model elaborations with simpler results, which requires a closed form, or in deducing asymptotic behavior, which requires at least some simple underlying structure. One interesting direction, not considered here, is to what extent the second class of matrices could "adequately" represent a more complicated covariance structure in providing forecasts. It will also be interesting to see whether more general matrix structures arise in practice, and are easily solved by methods not considered here.

REFERENCES

[1] Bodewig, E. Matrix Calculus. Second Edition, North-Holland, Amsterdam, 452 pp. (1959).

[2] Box, G.E.P. and Jenkins, G.M. Times Series Analysis, Forecasting and Control. Holden, Day, San Francisco (1970).

[3] Bühlmann, H. "Experience Rating and Credibility". ASTIN Bulletin, Vol. 4, Part 3, pp. 199-207 (July, 1967).

[4] Bühlmann, H. "Credibility Procedures". Sixth Berkeley Symposium on Mathematical Statistics, pp. 515-525 (1971).

[5] Bühlmann, H. and Straub, E. "Glaubwürdigkeit für Schadenssätze". Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker, Vol. 70, pp. 111-133 (1970).
Translation by C.E. Brooks. "Credibility for Loss Ratios". ARCH 1972.2.

[6] Gerber, H.U. and Jones, D.A. "Credibility Formulas of the Updating Type". Department of Mathematics, University of Michigan (1974).

[7] Jewell, W.S. "Multi-Dimensional Credibility". ORC 73-7, Operations Research Center, University of California, Berkeley (April, 1973).
To appear in Journal of Risk and Insurance.

[8] Jewell, W.S. "Credible Means are Exact Byaesian for Simple Exponential Families". ASTIN Bulletin, Vol. 8, Part 1, pp. 77-90 (September, 1974).

[9] Jewell, W.S. "Regularity Conditions for Exact Credibility". ORC 74-22, Operations Research Center, University of California, Berkeley (July, 1974).
To appear in ASTIN Bulletin.

[10] Jewell, W.S. "Exact Multi-Dimensional Credibility". ORC 74-14, Operations Research Center, University of California, Berkeley (May, 1974). Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker, Vol. 74, 2, pp. 193-214 (1974).

[11] Jewell, W.S. "Model Variations in Credibility Theory". ORC 74-25, Operations Research Center, University of California, Berkeley (August, 1974). To appear in Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, September, 1974.

[12] Kailath, T. "Fredholm Resolvents, Wiener-Hopf Equations, and Ricatti Differential Equations". IEEE Trans. on Info. Thy., Vol. 15, 6, pp. 665-672 (November, 1969).

[13] Kailath, T. "The Innovations Approach to Detection and Estimation Theory". Proc. IEEE, Vol. 58, pp. 680-695 (May, 1970).

[14] Lindley, D.V. Bayesian Statistics: A Review. SIAM, Philadelphia, 83 pp. (1972).

[15] Lindley, D.V. and Smith, A.F.M. "Bayes Estimates for the Linear Model". Journal Royal Statistical Society (B), Vol. 34, pp. 1-41 (1972).

[16] Roy, S.N. and Sarhan, A.E. "On Inverting a Class of Patterned Matrices". Biometrika, Vol. 43, Parts 1 & 2, pp. 227-231 (June, 1956).

[17] Tocher, K.D. "Discussion on M. Box and Dr. Wilson's Paper". Journal Royal Statistical Society (B), Vol. 13, pp. 39-42 (1951).

[18] Winkler, R.L. and Barry, C.B. "Nonstationary Means in a Multinormal Process". RR-73-9, IIASA, Laxenburg, Austria (October, 1973).

[19] Zellner, A. An Introduction to Bayesian Inference in Econometrics. J. Wiley & Sons, New York (1971).